Location History

P.I: Venu Garikapati
National Renewable Energy Laboratory
Presentation Date: June 22, 2021

DOE Vehicle Technologies Program
2021 Annual Merit Review and Peer Evaluation Meeting

Project ID # VAN041

This presentation does not contain any proprietary, confidential, or otherwise restricted information.

Overview

Timeline

- Project start date: 06/01/2020
- Project end date: 06/30/2021*
- Percent complete: 85%

* The analysis project will conclude by June 30. There *might* be an add-on effort funded through EEMS

Barriers

- Perennial data sources, and modeling techniques for understanding the complex role of the human decision-making process in mobility systems

Budget

- Total project funding: $120K
  - DOE share: $120K
  - Contractor share: $0
- Funding for FY 2020: $120K

Partners

- We intend to run the comparison of ML and traditional choice models by the travel modeling lead at the Colorado Department of Transportation

Relevance

- In **traditional travel surveys**, respondents **manually provide information** (purpose, mode, departure time etc.,) on each trip they made on a given day. This can prove cumbersome to the respondents

- Through mobile phone-based **location history (LH) data**, it is **possible to collect information on all the places** visited **and modes** used by an individual on a specific day

- This project aims to explore **location history data as an alternative to traditional travel surveys** for modeling individual travel patterns

- Also, in this project we will **compare machine learning (ML)** based modeling methods with **traditional choice models** for their efficacy in predicting activity time, mode, and departure time choice decisions

Learnings from this project can enhance our understanding of human decision-making process and advance the state of modeling in agent-based travel models.

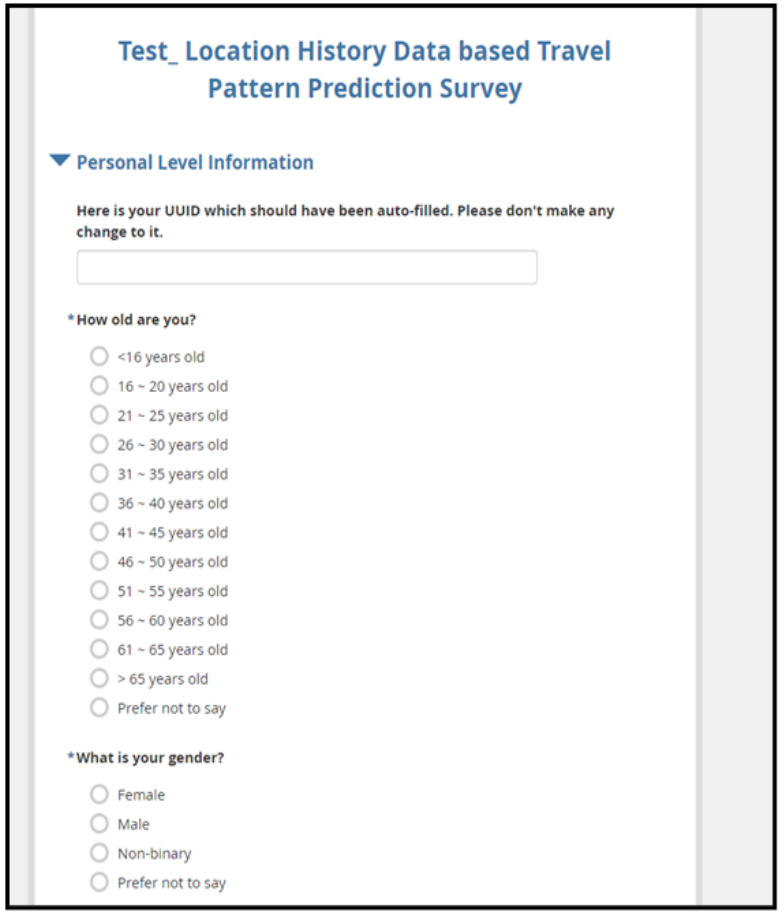
Approach

- Collect **location history (LH) data**, along with socioeconomic and demographic characteristics, from a **small sample of individuals**

- Using LH data, develop traditional choice models as well as ML models to predict
  - **what** activity will an individual participate in next
  - **when** will the individual start their journey to reach the next activity
  - **which** mode to will they use to get to that activity

- End goals
  - Assessment of **viability of location history (LH) for travel pattern prediction**
  - **Compare accuracies** of traditional choice models and machine learning models in predicting travel behavior

Approach: Data Collection Using E-mission



Test_ Location History Data based Travel Pattern Prediction Survey

Personal Level Information

Here is your UUID which should have been auto-filled. Please don't make any change to it.

*How old are you?

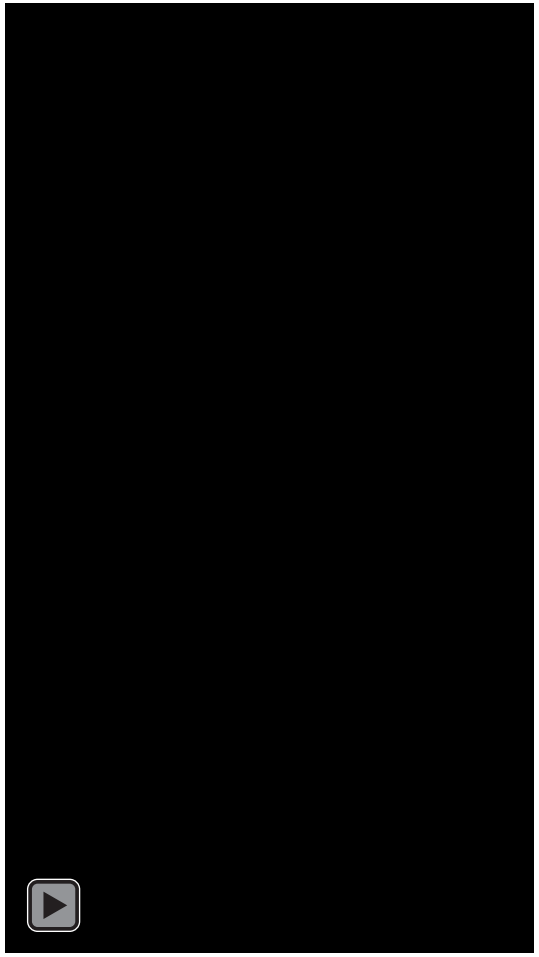
- <16 years old
- 16 ~ 20 years old
- 21 ~ 25 years old
- 26 ~ 30 years old
- 31 ~ 35 years old
- 36 ~ 40 years old
- 41 ~ 45 years old
- 46 ~ 50 years old
- 51 ~ 55 years old
- 56 ~ 60 years old
- 61 ~ 65 years old
- > 65 years old
- Prefer not to say

*What is your gender?

- Female
- Male
- Non-binary
- Prefer not to say

Socio-demographic survey

App-based Data Collection



Click to play video

Technical Accomplishments and Progress

The team resorted to analyzing a small fallback dataset which allowed the project to move forward.

Collect location history from a small sample of 50-100 individuals

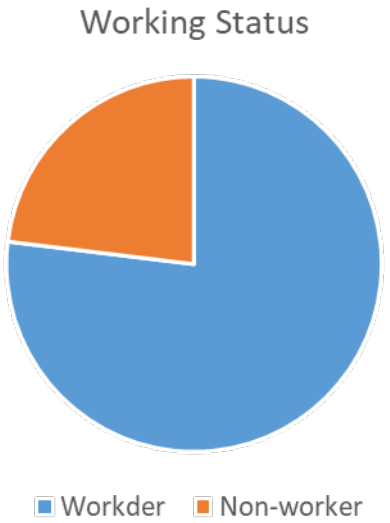
- Internal permissions took longer than anticipated due to COVID-19

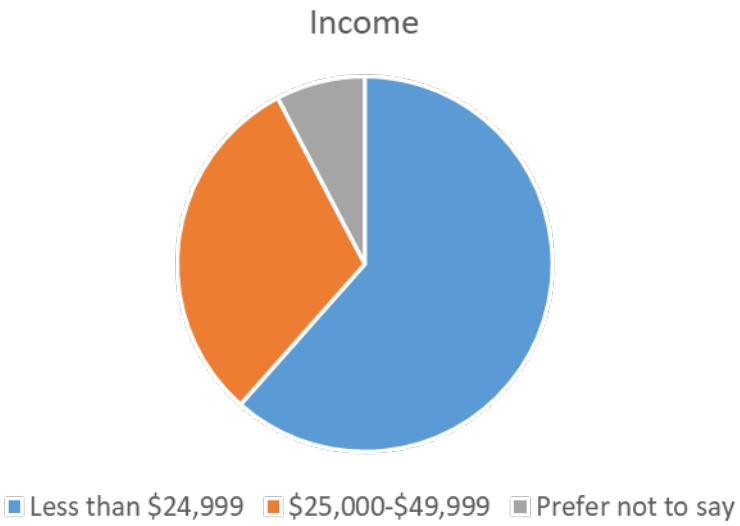
Other data collected through e-mission

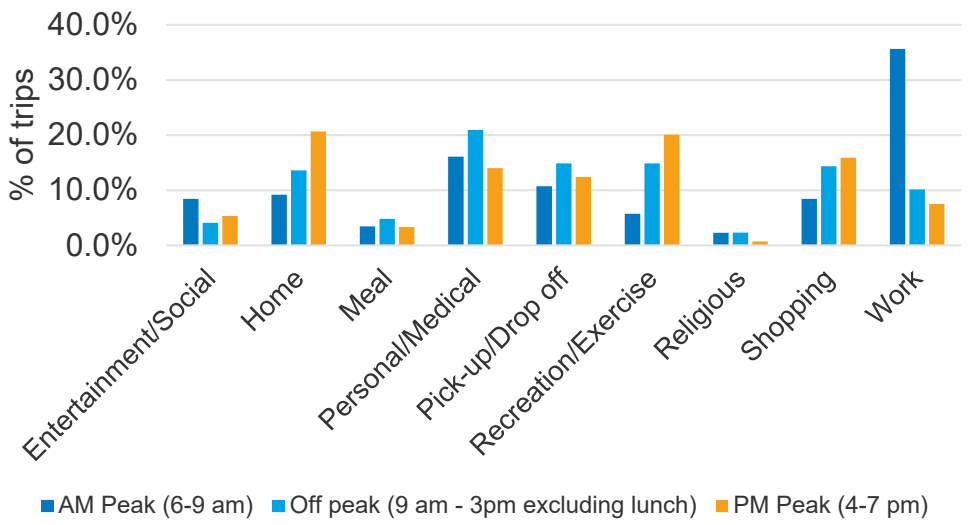
- Publicly accessible
- Collected for specific projects
  - UC Berkeley
  - University of New South Wales
  - **Colorado Energy Office** (CanBikeCo Survey)

Technical Accomplishments and Progress: CanBikeCo Survey

- Pre-pilot of the CanBikeCo program was **initiated by the Colorado Energy office**
  - CanBikeCo **smart phone application** (a version of the e-mission app being used in this project) + **sociodemographic survey**
- 13 participants provided (intermittent) data from **November 2020 to January 2021**
  - Higher proportion of low income, and working cohort
- 4163 raw trips were collected, of which **2,428 trips have complete information**







Technical Accomplishments and Progress: Results of Traditional Choice Model

* Preliminary Results

Multinomial Logit Model (MNL)

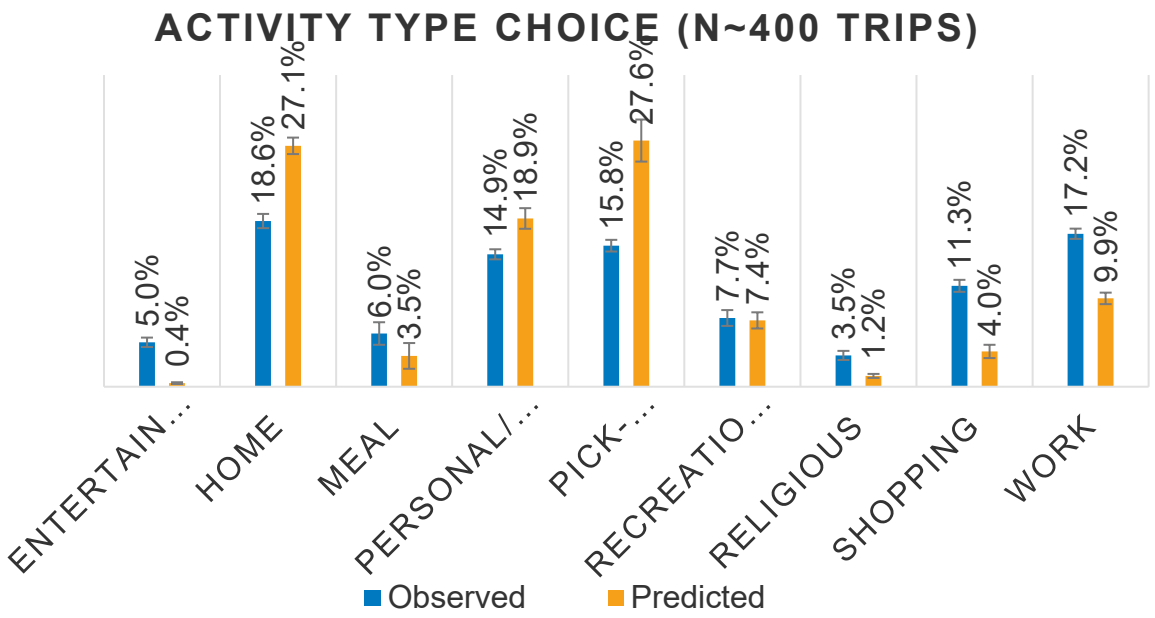
- Utility Function of MNL

$$V_i = \alpha + \sum_{j=1}^{J} \beta_j x_j + \varepsilon_i$$

- The probability of choosing alternative *i*

$$P(i|C) = \frac{e^{V_i}}{\sum_{k \epsilon C} e^{V_k}}$$

Where
i is one of the alternatives
α is the constant value;
J is the total number of alternative attributes considered
β_j is the parameter value of attribute x_j
C is the choice set



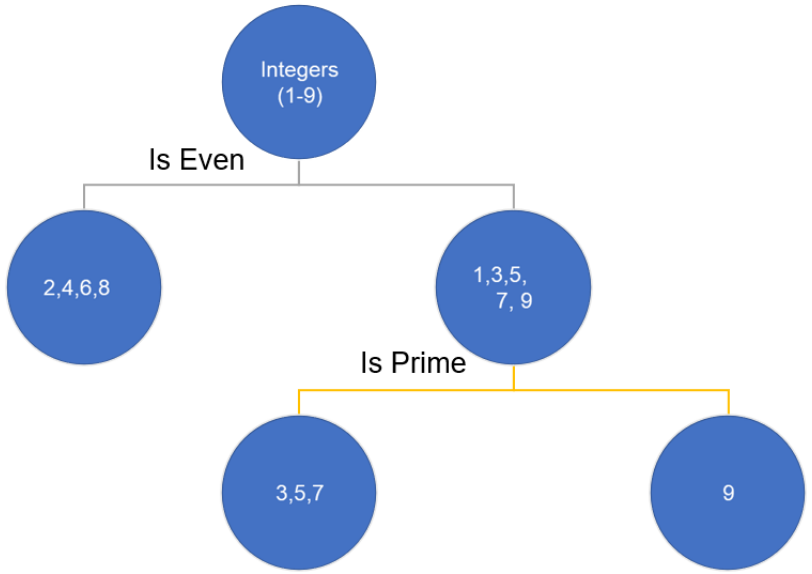
- Sample wise prediction accuracy: **39.7%**
- Important features in the model
  - Work from home
  - Employment Status
  - Adopted mode (walking, driving)
  - Am Peak (Time-of-day)

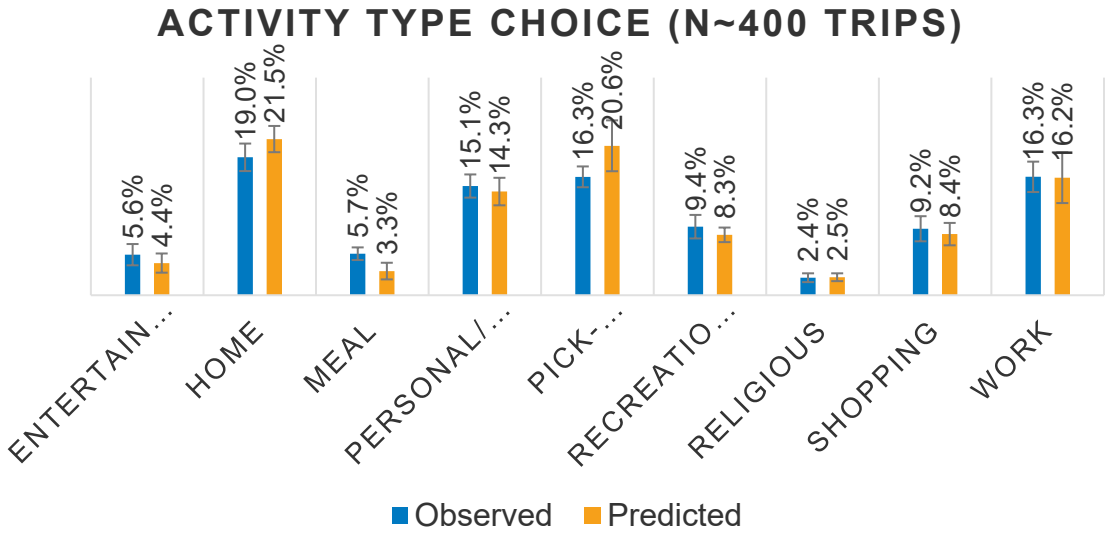
Model performance: **Inferior**

Technical Accomplishments and Progress: Machine Learning Based Model

* Preliminary Results

- Random Forest
  - Each tree models a series of random splits for each feature in the sample data
  - Estimation occurs by traversing the tree from root (top) to leaf (bottom) where a prediction is stored
  - Ensemble of such trees: all predictions are averaged





- Sample wise prediction accuracy: **50.46%**
- Important features in the model
  - Age
  - Count of all points of interest within 500 meters from the destination
  - Work from home
  - Time-of-day

Model performance: **Good**

Collaborations and Coordination with Other Institutions

- Results of the modeling exercise will be shared with travel modelers at the **Colorado Department of Transportation** for their feedback

- **Colorado energy office (CEO)** is moving forward with the next phase of the CanBikeCo project, in which e-mission will be used for data collection. We plan to request CEO for access to the data for enhancing the activity, mode, and departure choice models

Proposed Future Research

- Finish estimation of traditional and machine learning models for **activity type, mode, and departure time choice** using the small sample CanBikeCo dataset

- Attempt **native data collection for 50-100 individuals** using e-mission app
  - All the necessary permissions for data collection have been obtained now

- Enhance model estimation using **larger CEO dataset** (pending permission from CEO)

Any proposed future work is subject to change based on funding levels.

Summary

Objective

- Assessment of **viability of location history (LH) data** (along with identifying any shortcomings) **for travel pattern prediction**

Efforts over the past year focused on

- Groundwork for **native data collection** (delayed due to COVID-19)
- Analyzing **CanBikeCo data** for use in this project

Future efforts will focus on

- Finish **model estimation using the small sample** CanBikeCo data
- Attempt native data collection (pending available funding) and develop models using the data

Thank You

www.nrel.gov

NREL/PR-5400-80016

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Vehicle Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

NREL
Transforming ENERGY

Technical Back-Up Slides

Background

- **Four stage travel models** represent travel as **aggregate flows** (or trips) between origins and destinations, thus failing to capture the nuances in travel preferences at the level of an individual

- **Activity-based models or ABMs** (which are a form of agent-based models) represent travel at the level of an individual and **mimic the decision process that individuals follow** in making activity, location, mode, and route decisions

- **Shortcoming**: Even the most advanced ABMs rely on traditional survey data sources such as **travel diaries** to develop behavioral models.
  - Surveys are done at **infrequent intervals**; and are **limited in sample size**

Traditional vs. Machine Learning Models

Traditional Choice Model	Machine Learning Model
Requires assumptions on error terms used in model estimation	Does not require knowledge/assumptions of input data distributions
Have closed form solutions, and easily interpretable results	Back box estimation; Although feature importance can be derived, hard to assess the impact (directionality) of a variable on model prediction
Moderate computational requirements; a bit difficult to obtain replicable results	Low computational requirements (fast iteration, easy to replicate)